

Published in final edited form as:

Conf Proc IEEE Eng Med Biol Soc. 2012 ; 2012: 2166–2169. doi:10.1109/EMBC.2012.6346390.

Measurement of Food Volume Based on Single 2-D Image without Conventional Camera Calibration*

Yaofeng Yue¹, Wenyan Jia^{2,3}, and Mingui Sun³

Mingui Sun: drsun@pitt.edu

¹Department of Electrical Engineering, University of Pittsburgh

²Department of Neurosurgery, University of Pittsburgh

³Department of Neurosurgery, Bioengineering, and Electrical Engineering, University of Pittsburgh, Pittsburgh, PA, 15213, USA

Abstract

Food portion size measurement combined with a database of calories and nutrients is important in the study of metabolic disorders such as obesity and diabetes. In this work, we present a convenient and accurate approach to the calculation of food volume by measuring several dimensions using a single 2-D image as the input. This approach does not require the conventional checkerboard based camera calibration since it is burdensome in practice. The only prior requirements of our approach are: 1) a circular container with a known size, such as a plate, a bowl or a cup, is present in the image, and 2) the picture is taken under a reasonable assumption that the camera is always held level with respect to its left and right sides and its lens is tilted down towards foods on the dining table. We show that, under these conditions, our approach provides a closed form solution to camera calibration, allowing convenient measurement of food portion size using digital pictures.

I. INTRODUCTION

Obesity has increased remarkably in the U.S. and many developed and developing countries over the past decades and the rate of obesity continues to grow [1][2]. To study and control this epidemic, it is critical to monitor food intake of people in their real lives. Conventional dietary assessment methods include food diary, 24-hour recall, and food frequency questionnaire (FFQ) [3]. However, since these methods are subjective, study participants often underreport their intake due to unwillingness to report accurately or incomplete recall from their memory.

Using a camera to document food has become feasible, thanks to the rapid development of mobile devices, such as smartphones, portable or wearable digital cameras, and tablet computers. Food images taken by these means can be used to not only help participants during their dietary recall process, but also measure food volumes. However, in the latter case, a reference object is required for camera calibration which establishes a correspondence between the coordinate systems of the camera and the real-world. Traditionally, this reference object is a black-white or a colored checkerboard card [5][6][7]. Although this card is effective for camera calibration, it is not convenient in practice because the study participant must place this card besides food when taking a picture.

*This work was supported by National Institutes of Health grant U01 HL91736

Another problem in image based food volume estimation is associated with camera calibration which is required to establish the correspondence between the camera and the real world. This procedure usually utilizes the knowledge of the focal length of the camera. However, when a camera has an auto-focus function, this value of the focal length is usually not provided to the user. In this work, we solve these problems by developing new references and algorithms. In practice, foods are often served using a circular container, such as a dining plate, a bowl, or a cup. We thus use the food container itself as a reference object [8][9]. In addition, we establish a new mathematical model that does not require explicit knowledge of the focal length.

II. METHODS

The well-known pin-hole camera model provides a perspective projection from the object plane to the image plane. Figure 1 illustrates the geometric relation between a circular reference in the object plane and its projected feature, an ellipse, in the image plane. In this figure, O is the optical center of the camera, f is the unknown focal length, d_0 is the z -coordinate of the intersection point of the object plane and the z -axis. Vertex point O and the radial lines from O passing through the boundary of the circular reference form a cone. In this section, a mathematical model for this cone will first be established under coordinate system xyz . To simplify the derivation, this coordinate system will be linearly transformed to a new coordinate system $x'y'z'$ such that the object plane will be parallel to the $x'y'$ plane. Then, the orientation of the object plane is solved using a simplified circular equation in $x'y'z'$. Finally, with the given diameter of the circular reference, key perspective projections between the object and image planes are established and user-selected food dimensions will be measured which lead to the estimation of food volume (portion size).

A. System Transformation

The perspective projection of a circular feature reference is always an ellipse [10], which has a general form of

$$a'x^2 + 2h'xy + b'y^2 + 2g'x + 2f'y + d' = 0 \quad (1)$$

where, a' , h' , b' , g' , f' and d' are coefficients, $d' \neq 0$. Since the focal length f and pixel size p_x are both unknown, we are unable to obtain the ellipse equation in xyz . In the image plane, if we denote (x_{pi}, y_{pi}) as the pixel coordinates of a point i on the ellipse boundary and (c_1, c_2) as the principle point of the image, then the coordinates of point i is

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} x_{ci} \\ y_{ci} \end{bmatrix} \cdot P_x \quad (2)$$

where $x_{ci} = x_{pi} - c_1$ and $y_{ci} = y_{pi} - c_2$. Assuming we have obtained n_p points from the ellipse boundary in the picture, from (1) we have

$$A \cdot P_x \cdot u = -d' \cdot \mathbf{1} \quad (3)$$

where

$$\mathbf{A} = \begin{bmatrix} x_{c1}^2 & 2x_{c1}y_{c1} & y_{c1}^2 & 2x_{c1} & 2y_{c1} \\ x_{c2}^2 & 2x_{c2}y_{c2} & y_{c2}^2 & 2x_{c2} & 2y_{c2} \\ x_{c3}^2 & 2x_{c3}y_{c3} & y_{c3}^2 & 2x_{c3} & 2y_{c3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{cnp}^2 & 2x_{cnp}y_{cnp} & y_{cnp}^2 & 2x_{cnp} & 2y_{cnp} \end{bmatrix} \quad (4)$$

$$\mathbf{P}_x = \begin{bmatrix} p_x^2 & & & & \\ & p_x^2 & & & \\ & & p_x^2 & & \\ & & & p_x & \\ & & & & p_x \end{bmatrix} \quad (5)$$

$$\mathbf{u} = [a' h' b' g' f']^T \text{ and } \mathbf{1} = [1 \ 1 \ 1 \ 1 \ 1]^T$$

Since many points can be selected from the ellipse boundary on the picture, $n_p = 5$. Solving for \mathbf{u} in (3) using the least square method, we obtain

$$\mathbf{u} = -\mathbf{d}' \cdot \mathbf{P}_x^{-1} \cdot (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \cdot \mathbf{1} \quad (6)$$

In (6), we denote

$$[t_1 \ t_2 \ t_3 \ t_4 \ t_5]^T = -(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \cdot \mathbf{1} \quad (7)$$

which can be directly calculated from the picture. Then

$$\mathbf{u} = [t_1 p_x^{-2} \ t_2 p_x^{-2} \ t_3 p_x^{-2} \ t_4 p_x^{-1} \ t_5 p_x^{-1}]^T \cdot \mathbf{d}' \quad (8)$$

As shown in [8] [9], the equation of the cone formed by the ellipse in the image plane and the camera center O as in Figure 1 is

$$[x \ y \ z] \cdot \mathbf{Q} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0 \quad (9)$$

where

$$\mathbf{Q} = \begin{bmatrix} a' f^2 & h' f^2 & g' f \\ h' f^2 & b' f^2 & f' f \\ g' f & f' f & d' \end{bmatrix} \quad (10)$$

If we denote the orientation of the object plane as (l, m, n) , the object plane is

$$lx + my + nz = nd_0 \quad (11)$$

where

$$l^2+m^2+n^2=1 \quad (12)$$

Then, a linear coordinate transformation (13) is conducted such that the object plane is parallel to $x'y'z'$ plane in the new coordinate system $x'y'z'$. In this work, we use a transformation matrix T in Eq. (14) [4].

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = T \cdot \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad (13)$$

$$T = \begin{bmatrix} -\frac{m}{\sqrt{l^2+m^2}} & -\frac{nl}{\sqrt{l^2+m^2}} & l \\ \frac{l}{\sqrt{l^2+m^2}} & -\frac{mn}{\sqrt{l^2+m^2}} & m \\ 0 & \sqrt{l^2+m^2} & n \end{bmatrix} \quad (14)$$

Note that when the ellipse in the image follows a standard circle equation, $l = m = 0$, $n = 1$. There is no need to transform the coordinate system.

Figure 2 illustrates the perspective projection under the new coordinate system $x'y'z'$. The object plane (11) becomes (15) after the transformation, which indicates the object plane is orthogonal to the z' axis.

$$z' = nd_0 \quad (15)$$

After the transformation, Eq. (9) becomes

$$\begin{bmatrix} x' & y' & z' \end{bmatrix} \cdot Q_1 \cdot \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = 0 \quad (16)$$

where

$$Q_1 = T^T \cdot Q \cdot T = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} \quad (17)$$

B. Orientation Estimation

In the new coordinate system $x'y'z'$, the object plane is parallel to $x'y'$ plane. If we substitute (15) into (16), it should reduce to a standard circle equation, i.e. $q_{11} = q_{22}$ and $q_{12} = q_{21} = 0$ in (17). Combining (10) and (14), it can be derived that

$$a' f^2 \frac{m^2}{\tau^2} + b' f^2 \frac{l^2}{\tau^2} - 2h' f^2 \frac{ml}{\tau^2} - a' f^2 \frac{l^2 n^2}{\tau^2} = b' f^2 \frac{m^2 n^2}{\tau^2} + d' \tau^2 + 2h' f^2 \frac{mln^2}{\tau^2} - 2g' fnl - 2f' fmn$$

and

$$(a' - b')f^2 \frac{lmn}{\tau^2} + h' f^2 \frac{m^2 n}{\tau^2} - h' f^2 \frac{nl^2}{\tau^2} - g' mf + f' lf = 0$$

where $\tau^2 = f^2 + m^2$.

Substituting (8) into above two equations and denoting $k = \frac{f}{\rho_x}$, we obtain

$$t_1 k^2 \frac{m^2}{\tau^2} + t_3 k^2 \frac{l^2}{\tau^2} - 2t_2 k^2 \frac{ml}{\tau^2} - t_1 k^2 \frac{l^2 n^2}{\tau^2} = t_3 k^2 \frac{m^2 n^2}{\tau^2} + \tau^2 + 2t_2 k^2 \frac{mn^2 l}{\tau^2} - 2t_4 knl - 2t_5 kmn \quad (18)$$

$$(t_1 - t_3)k \frac{mnl}{\tau^2} + t_2 k \frac{m^2 n}{\tau^2} - t_2 k \frac{nl^2}{\tau^2} - t_4 m + t_5 l = 0 \quad (19)$$

Therefore, three equations (12), (18) and (19) are available but there are four unknowns, l , m , n and k . Unless additional information is added to the system, the equations cannot be solved. In this work, we simplify the problem by considering a common practice when a person takes pictures. We assume that when the food pictures are taken, the bottom of the camera is always parallel to the dining table surface, while the lens is allowed to tilt towards the food. Mathematically, this condition is equivalent to

$$l = 0 \quad (20)$$

Then, we can solve for m , n

$$n = \sqrt{\frac{t_1 t_4^2}{t_3 t_4^2 - 2t_2 t_4 t_5 + t_2^2}} \quad (21)$$

$$m = \sqrt{1 - n^2} \quad (22)$$

Only the positive solution of n is retained because food is always in front of the camera. In addition, two possible solutions of m are obtained with different signs. In practice, respecting to the plane passing through the camera center and parallel to the dining table surface, one solution refers to the situation that the table surface is below it while the other solution above it. Here, we take the positive solution which reflects the reality.

Then, the fourth unknown, the camera parameter, is given by

$$k = \sqrt{\frac{(t_3 - t_1)t_4^2 + t_2^2 - 2t_2 t_4 t_5}{t_1 t_2^2}} \quad (23)$$

C. Dimension Estimation

Knowing the orientation (l, m, n) is not enough to perform dimensional measurements of the object because the distance between camera and object is not uniquely determined. In this work, we assume that the diameter of the physical circular feature, e.g., the diameter of the

dining plate, D , is given. With the solutions (20), (21), (22) and (23), the standard form of circle equation in the object plane is

$$\left(x' - \frac{e_2}{e_1}d_0\right)^2 + \left(y' + \frac{e_3}{e_1}d_0\right)^2 = \frac{e_2^2 + e_3^2 - e_4^2}{e_1^2}d_0^2 \quad (24)$$

where $e_1 = t_1k^2$, $e_2 = mnt_2k^2 + t_4kn^2$, $e_3 = mn^2 + t_5km^2n - t_3mk^2n^2 - kt_5n^3$ and $e_4 = (t_3k^2n^2 + n^2 + 2t_5kmm)n^2$.

Then, d_0 can be obtained as

$$d_0 = \frac{D}{2} \sqrt{\frac{e_2^2 + e_3^2 - e_4^2}{e_1^2}} \quad (25)$$

Therefore, the distance from camera center O to the object plane is determined, i.e. nd_0 . Then, object dimensions in or related to the object plane can be estimated. For food portion size estimation, two types of dimensions are especially important, DE and EF as illustrated in Figure 3, where DE is in the object plane and EF is orthogonal to it with point E in the plane. For DE in Figure 3(a), we denote the pixel coordinates of the two points in the image as (x_{p1}, y_{p1}) and (x_{p2}, y_{p2}) . For EF in Figure 3(b), we denote the pixel coordinates of the two points as (x_{p3}, y_{p3}) and (x_{p4}, y_{p4}) . Then, it has been shown in [8][9] that the two dimensions can be calculated as

$$|DE| = nd_0 \sqrt{g_1^2 + g_2^2} \quad (26)$$

where, $g_1 = \frac{x_{c2}}{my_{c2} + nk} - \frac{x_{c1}}{my_{c1} + nk}$ and $g_2 = \frac{mk - ny_{c2}}{my_{c2} + nk} - \frac{mk - ny_{c1}}{my_{c1} + nk}$ and

$$|EF| = nd_0 \left(1 - \frac{m(y_{p4} - c_2) + nk}{m(y_{p3} - c_2) + nk} \cdot \frac{n(y_{p3} - c_2) + mk}{n(y_{p4} - c_2) + mk}\right) \quad (27)$$

where $x_{ci} = x_{pi} - c_1$ and $y_{ci} = y_{pi} - c_2$.

III. EXPERIMENTS

In our experiments, lengths and thicknesses of six food replicas, including roll, strawberry, cake, bread, chicken, and carrot, were calculated using a dining plate of diameter 26cm as the reference. The plate was placed randomly on a table at different positions, but was always kept observable when pictures were taken using hand held digital cameras. As stated early, the cameras were positioned approximately level with respect to the table surface while the lens was tilted down to obtain satisfactory food pictures. The viewing angles were randomly distributed between 20° and 80°. Figure 4 shows three typical pictures. Two digital cameras were used, both having the auto-focus function. The length and thickness of the food were selected manually from the displayed food image on the computer screen using a mouse. For some irregularly shaped foods, such as chicken legs and carrot pieces, the thickness of the food was hard to determine because of the lack of edges on top surfaces of the food, and hence only lengths were measured. Then, the plate boundary, an ellipse, in the image was automatically detected [11] and was used to calculate the orientation of the table surface using Eqs. (7), (20), (21) and (22). With the known diameter of the plate, the

dimensions were then calculated using (26) and (27). Table I provides more detailed estimation results.

It can be observed that only the thicknesses of the roll and strawberry have an average error more than five percent (in terms of absolute values). The major cause of the error was due to the selection of a pair of points which represent the thickness of the food. In practice, however, this pair of points was not easy to determine. The difficulty increased proportionally with the irregularity in food shape. Nevertheless, our approach provides reasonable results for most food items provided that the target dimensions are properly represented on the image.

IV. CONCLUSION

We have presented a convenient approach to the estimation of food volume based on a single 2-D food image as the input without the need of using a conventional camera calibration procedure. Since critical food dimensions used for volume calculation are solved with closed-form solutions, this algorithm can be easily implemented as an application program in smartphones. The program interface may allow users to use food pictures to estimate food volumes which can further be linked to a food database to estimate calories and nutrients.

References

1. Basu A. Forecasting distribution of body mass index in the United States: Is there more room for growth. *Medical Decision Making*. 2010; 30(3)
2. Wang YC, Colditz GA, Kuntz KM. Forecasting the obesity epidemic in the aging U.S. population. *Obesity*. 2007; 15(11):2855–2865. [PubMed: 18070778]
3. Thompson, FE.; Subar, AF. *Nutrition in the Prevention and Treatment of Disease*. 2. 2008. Dietary assessment methodology.
4. Safaee-Rad R, Tchoukanov I, Smith KC, Benhabib B. Three-Dimensional location estimation of circular features for machine vision. *IEEE Trans Robotics Automat*. Oct; 1992 8(5):624–640.
5. Sun M, Liu Q, Schmidt K, Yang J, Yao N, Fernstrom JD, Fernstrom MH, Delany JP, Sclabassi RJ. Determination of food portion size by image processing. *Conf Proc IEEE Eng Med Biol Soc*. Aug. 2008 :871–874. [PubMed: 19162795]
6. Boushey CJ, Kerr DA, Wright J, Lutes KD, Ebert DS, Delp EJ. Use of technology in children's dietary assessment. *European Journal of Clinical Nutrition*. 2009; 63:S50–S57. [PubMed: 19190645]
7. Weiss R, Stumbo PJ, Divakaran A. Automatic food documentation and volume computation using digital imaging and electronic transmission. *J Am Diet Assoc*. 2010; 110:42–44. [PubMed: 20102824]
8. Jia W, Yue Y, Fernstrom JD, Yao N, Sclabassi RJ, Fernstrom MH, Sun M. Image based estimation of food volume using circular referents in dietary assessment. *Journal of Food Engineering*. Mar; 2012 109(1):76–86. [PubMed: 22523440]
9. Yue Y, Jia W, Fernstrom JD, Sclabassi RJ, Fernstrom MH, Yao N, Sun M. Food volume estimation using a circular reference in image-based dietary studies. *Bioengineering Conference, Proceedings of the 2010 IEEE 36th Annual Northeast*. Mar.2010 :1–2.
10. Narayan, Shanti. *Analytical Solid Geometry*. 12. S. Chand And Company; 1961.
11. Nie J, Wei Z, Jia W, Li L, Fernstrom JD, Sclabassi RJ, Sun M. Automatic detection of dining plates for image-based dietary evaluation. *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. :4312–4315.Nov.2010

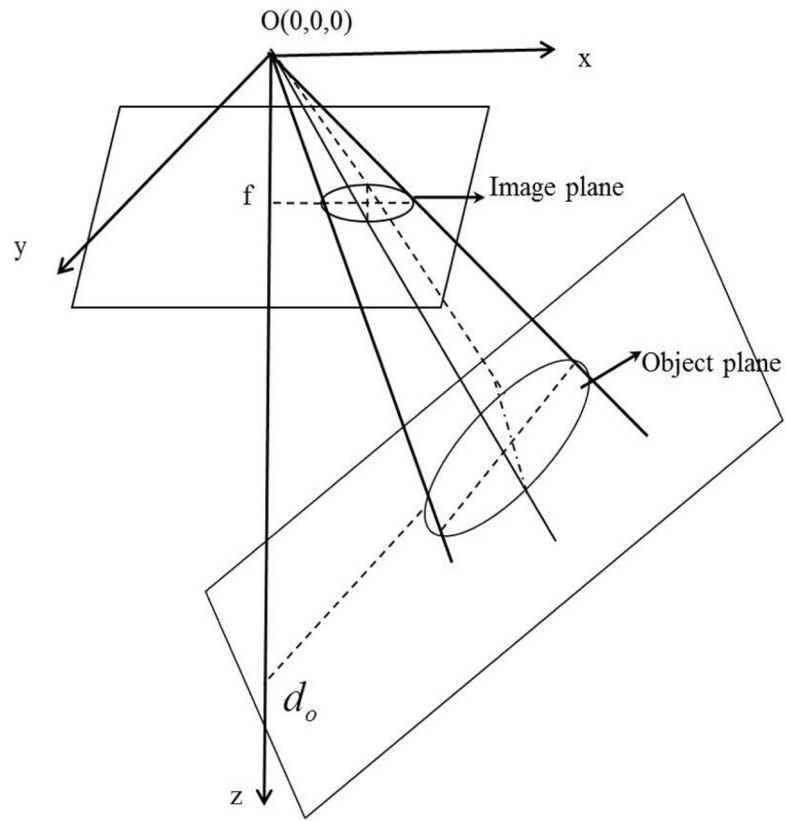


Fig. 1. Projective system of the optical center, image and object in coordinate system xyz .

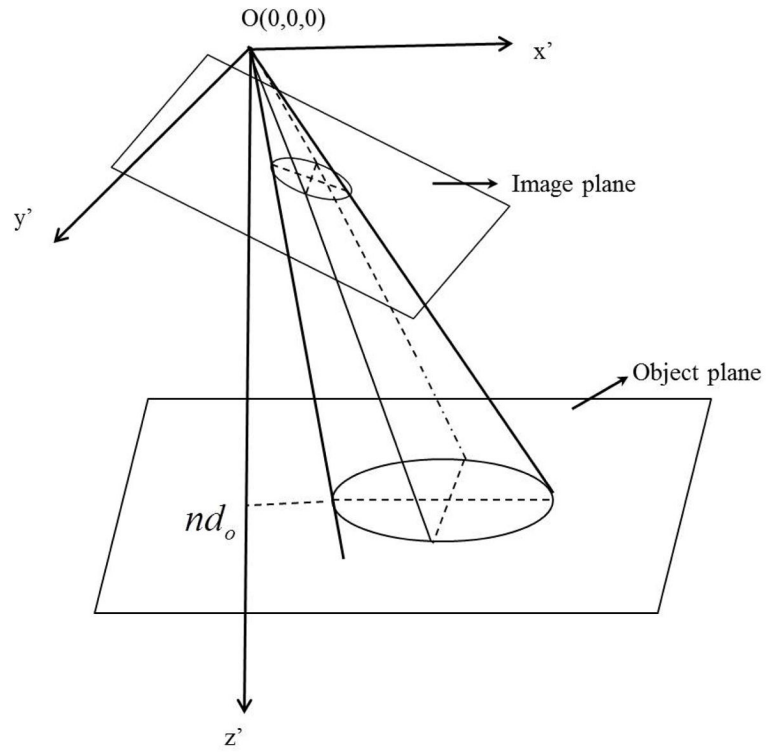


Fig. 2. Schematic diagram of the projective system in new coordinate system $x'y'z'$ after linear transformation (13).

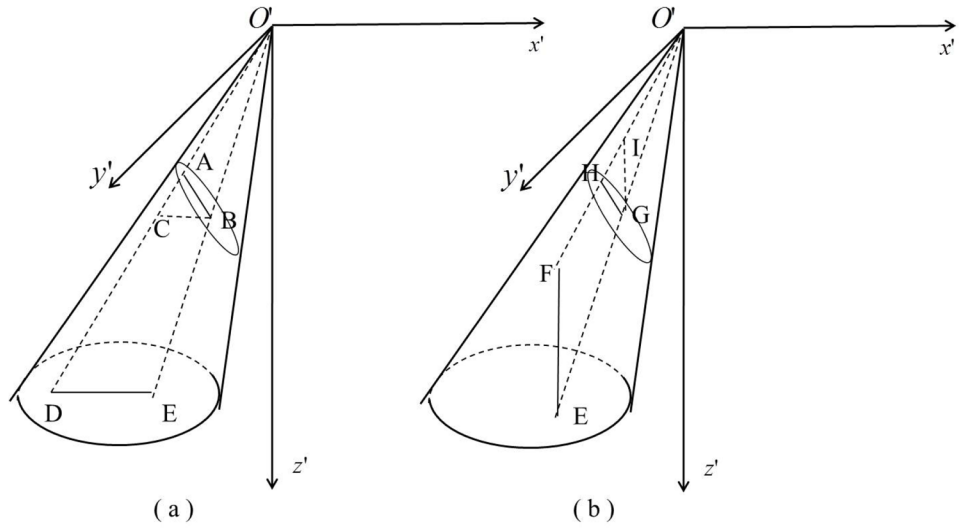


Fig. 3.
 (a) Dimension DE in the object plane. (b) Dimension EF perpendicular to the object plane.



Fig. 4.
Typical food pictures. Plates of foods were randomly placed on the table.

TABLE I

Estimation results of the selected food dimensions.

Food	Dim. Type	Pictures	Truth (mm)	Mean (mm)	Std. (mm)	Error (%)
Roll	Length	20	70	68.63	2.45	-1.95
	Thickness	20	42	46.65	2.63	11.08
Strawberry	Length	24	28	27.44	1.69	-2.01
	Thickness	24	30	32.22	5.55	7.38
Cake	Length	26	48	46.98	1.79	-2.13
	Thickness	26	38	37.56	1.58	-1.16
Bread	Length	26	100	96.03	4.09	-3.97
	Thickness	26	13	13.19	1.07	1.49
Chicken	Length	20	110	113.17	4.46	2.88
Carrot	Length	24	108	108.05	4.50	0.05